

UNITED STATES PATENT APPLICATION

**SCALABLE FLASH/NV STRUCTURES AND DEVICES  
WITH EXTENDED ENDURANCE**

**INVENTOR**

Arup Bhattacharyya

Schwegman, Lundberg, Woessner, & Kluth, P.A.  
1600 TCF Tower  
121 South Eighth Street  
Minneapolis, Minnesota 55402  
ATTORNEY DOCKET 1303.023US1  
MICRON 01-0565

094495 05674650

## SCALABLE FLASH/NV STRUCTURES AND DEVICES WITH EXTENDED ENDURANCE

### Technical Field of the Invention

5           This invention relates generally to integrated circuits and, more particularly, to nonvolatile programmable memory cells and reprogrammable logic circuit elements.

### Background of the Invention

10           General CMOS silicon gate technology has been scaled rapidly from the 1.0  $\mu\text{m}$  generation ( $V_{\text{DD}} = 5.0 \text{ V}$ ) to the 0.13  $\mu\text{m}$  generation ( $V_{\text{DD}} = 1.5 \text{ V}$ ) over a period of little over a decade. However, the progress of scaling Flash/NV programming voltage and power has been limited. Additionally, the Flash/NV devices have a limited endurance. That is, these devices are capable of performing a limited  
15   number of write/erase cycles that prevent their use in a number of applications. The progress of extending the endurance of Flash/NV devices also has been limited.

Most NV devices have a floating gate and use a power hungry channel hot electron write process and a tunnel erase process through the tunnel oxide. The tunnel erase process requires both a high programming voltage and a programming  
20   field that approaches the breakdown of the oxide. Thus, conventional Flash/NV devices require attributes of high voltage technology and circuitry in an environment of scaled low voltage CMOS technology. As such, the integration of NV devices with general high performance logic technology and DRAM technology is complicated. Additionally, the endurance of these NV devices is limited to about  
25    $10^5$ - $10^6$  cycles. Therefore, providing embedded Flash/EPROM in a general random logic or DRAM environment to achieve enriched functions requires complex circuitry and many additional masks, resulting in a relatively limited yield and high cost.

Two phase insulating materials, referred to as silicon-rich insulators (SRI), are known. SRI includes controlled and fine dispersions of crystalline silicon in a background of stoichiometric insulator such as  $\text{SiO}_2$  (referred to as silicon-rich oxide or SRO), or  $\text{Si}_3\text{N}_4$  (referred to as silicon-rich nitride or SRN). A unique set of  
5 insulators with a controlled and wide range of electrical properties can be formed by controlling the amount, distribution and size of silicon crystals.

SRI materials are capable of possessing charge trapping and charge injecting properties. "Charge-centered" SRI has refractive index in a range that provides the material with the property for trapping electrons or holes at the silicon centers due to  
10 the creation of quantum potential wells. "Injector" SRI has a refractive index in a range that provides the material with silicon centers that are within tunneling distance of each other such that charge can readily communicate between the charge centers. Injector SRI is characterized by high conductivity and behaves like semi-metal. The apparent high frequency dielectric constant of this material is greater  
15 than that of silicon. When superimposed on top of a dielectric, charge injected into this material from a metal plate is uniformly distributed to the silicon centers, which in turn injects charges uniformly into the insulator when biased. Thus, the injector SRI reduces local field fluctuations due to defects. At the same time, a large number of silicon injector centers at the insulator interface provides a geometrical  
20 pattern that enhances the tunneling, and thus the charge transfer or conduction, at significantly reduced average fields. This material has been termed an "injector" because of this enhanced tunneling.

Charge-centered or trapping SRI is a charge storing medium that includes charge-centered SRO characterized by a refractive index in the approximate range of  
25 1.5 to 1.6, and further includes charge-centered SRN characterized by a refractive index in the approximate range from 2.1 to 2.2. Injector SRI is a charge injector medium that includes injector SRO characterized by a refractive index that is approximately 1.85 and greater, and further includes injector SRN characterized by

a refractive index that is approximately 2.5 and greater. It was observed that the SRN class of materials was significantly more stable at high temperature compared to SRO in terms of interdiffusion and growth of silicon centers during high temperature processing as well as in terms of providing a reproducible interface  
5 between silicon and SRI and/or SiO<sub>2</sub> and SRI.

It has been proposed to use charge centered and injector SRI material in a variety of NV FET structures and associated Flash, PROM, EPROM, EEPROM, antifuse cells and arrays. In one of these proposed embodiments, the gate insulator stack includes a tunnel oxide, a thin layer of charge-centered SRN to trap charges  
10 and thereby act like a "floating plate," an overlayer of thicker CVD oxide, and a layer of injector SRN. The top CVD oxide is designed to prevent charge loss at the operating field and to be optimized for the appropriate programming voltage. The equivalent oxide thickness ( $t_{ox}$ ) of the gate insulator stack is primarily dependent on the tunnel oxide and the barrier CVD oxide thickness. The stack is scalable with  
15 respect to the programming voltage because the required programming field is reduced to 6-7E6 V/cm due to the injector-induced enhanced tunneling compared to 10-11E6 V/cm for a conventional NV/FET structure. Additionally, the programming gate voltage is directly coupled into the charge-centered layer to provide 100% coupling efficiency compared to the typical floating gate structures  
20 where capacitor divider effects and the cell geometry determine the coupling efficiency. The coupling efficiency for a floating gate structure is typically around 50%-70%. These proposed devices were shown to exhibit many orders of magnitude greater retention because of the reduced programming field. These devices are significantly more power efficient as they are written to and erased by  
25 direct tunneling, rather than by channel hot electron injection. However, the write/erase fields were still too high, and both the endurance and power reductions were still too limited.

Silicon "quantum dots" of 3 nm to 10 nm diameter have been fabricated in a controlled manner by either Low Pressure Chemical Vapor Deposition (LPCVD) followed by oxidation or by gas phase pyrolysis of silane to create nano crystal silicon aerosol. It has been proposed to either place these silicon nano crystals on top of the tunnel oxide or embed them into the gate insulator oxide. These nano crystals behave as charge centers similar to the charge-centered SRI layer described above. NV FET gate stacks were formed with the silicon quantum dots by adding a thicker oxide overlayer. While somewhat reduced voltage write/erase and up to 10E6 endurance were demonstrated, the write/erase fields were still too high, and both the endurance and the power reductions were still too limited.

Therefore, there is a need in the art to provide Flash/NV technology that overcomes these problems by being capable of using scalable programmable voltages and power, by being easily integratable with general scaled logic technology while minimizing the overhead associated with Flash/NV technology features, by extending endurance, and by providing faster write-erase cycles without impacting retention and reliability.

### Summary of the Invention

The above mentioned problems are addressed by the present subject matter and will be understood by reading and studying the following specification. The present subject matter provides a scalable Flash/NV structure that further extends the scalability of NV technology by providing a gate stack with a high K dielectric, a charge center or charge storing medium and at least one charge injector medium. The present subject matter requires a lower programming field, improves endurance, achieves faster write-erase cycles, and has variables that are capable of being manipulated for scaling purposes.

One aspect is a gate stack for a nonvolatile device. According to one embodiment, the gate stack includes a tunnel medium, a high K charge blocking and

charge storing medium, and an injector medium. According to this embodiment, the high K charge blocking and charge storing medium is disposed on the tunnel medium. Also according to this embodiment, the injector medium is operably disposed with respect to the tunnel medium and the high K charge blocking and charge storing medium to provide charge transport by enhanced tunneling. According to one embodiment, the injector medium is disposed on the high K charge blocking and charge storing medium. According to one embodiment, the tunnel medium is disposed on the injector medium.

One embodiment of the gate stack includes a first injector medium, a tunnel medium disposed on the first injector medium, a high K charge blocking and charge storing medium disposed on the tunnel medium, and a second injector medium disposed on the high K charge blocking and charge storing medium.

According to one embodiment, the gate stack includes a tunnel medium, a high K charge storing medium disposed on the tunnel medium, a high K charge blocking medium stored on the high K charge storing medium, and an injector medium. The injector medium is operably disposed with respect to the tunnel medium, the high K charge storing medium and the high K charge blocking medium to provide charge transport by enhanced tunneling. According to one embodiment, the injector medium is disposed on the high K charge blocking and charge storing medium. According to one embodiment, the tunnel medium is disposed on the injector medium.

One embodiment of the gate stack includes a first injector medium disposed on a substrate, a tunnel medium disposed on the first injector medium, a high K charge storing medium disposed on the tunnel medium, a high K charge blocking medium stored on the high K charge storing medium, and a second injector medium disposed on the high K charge blocking medium.

These and other aspects, embodiments, advantages, and features will become apparent from the following description of the invention and the referenced drawings.

5                                    **Brief Description of the Drawings**

Figure 1 is a graph showing refractive index of silicon-rich silicon nitride films versus  $\text{SiH}_2\text{Cl}_2/\text{NH}_3$  flow rate ratio.

Figure 2 is a graph showing current density versus applied field for silicon-rich silicon nitride films having different percentages of excess silicon.

10            Figure 3 is a graph showing flat band shift versus time at an applied field of  $4 \times 10^6$  volts/cm for silicon-rich silicon nitride films having varying percentages of excess silicon.

15            Figure 4 is a graph showing flat band shift versus time at an applied field of  $7 \times 10^6$  volts/cm for silicon-rich silicon nitride films having varying percentages of excess silicon.

Figure 5 is a graph showing apparent dielectric constant K versus refractive index for both Silicon Rich Nitride (SRN) and Silicon Rich Oxide (SRO).

Figure 6 is a cross-section view of a conventional nonvolatile field effect transistor (NV FET) device or Flash device.

20            Figure 7 illustrates the capacitive coupling for a conventional Flash device.

Figure 8 illustrates the capacitive coupling for a nonvolatile floating plate device.

Figure 9 illustrates the average field enhancement due to the incorporation of a top injection layer in a gate stack for a nonvolatile floating plate device.

25            Figure 10 illustrates the average field enhancement due to the incorporation of a bottom injection layer in a gate stack for a nonvolatile floating plate device.

Figure 11 illustrates the average field enhancement due to the incorporation of both a top injection layer and a bottom injection layer in a gate stack for a nonvolatile floating plate device.

Figure 12 illustrates the average field enhancement due to the incorporation of a high K dielectric and a top injection layer in a gate stack for a nonvolatile floating plate device.

Figure 13 illustrates the average field enhancement due to the incorporation of a high K dielectric and a bottom injection layer in a gate stack for a nonvolatile floating plate device.

Figure 14 illustrates the average field enhancement due to the incorporation of a high K dielectric and both a top injection layer and a bottom injection layer in a gate stack for a nonvolatile floating plate device.

Figure 15 is a graph showing floating charge versus average programming field for programming voltages applied to a floating gate of a conventional flash device (VP1), a NV floating plate device (VP2), a NV floating device having a gate stack formed with a single injection layer (VP3), and a NV floating device having a gate stack formed with a high K dielectric and a single injection layer (VP4).

Figure 16 is a graph showing the relationship between the log of the write/erase cycle, or endurance, and the average programming field.

Figure 17 is one embodiment of a NV floating plate device.

Figure 18 is one embodiment of a single injector layer gate stack for the NV floating plate device of Figure 17.

Figure 19 is one embodiment of a single injector layer gate stack for the NV floating plate device of Figure 17.

Figure 20 is one embodiment of the single injector layer gate stack of Figure 18.

Figure 21 is one embodiment of the single injector layer gate stack of Figure 18.



Figure 22 is one embodiment of a single injector layer gate stack for the NV floating plate device of Figure 17.

Figure 23 is one embodiment of a single injector layer gate stack for the NV floating plate device of Figure 17.

5        Figure 24 is one embodiment of the single injector layer gate stack of Figure 23.

Figure 25 is one embodiment of a double injector layer gate stack for the NV floating plate device of Figure 17.

10       Figure 26 is one embodiment of the double injector layer gate stack of Figure 25.

Figure 27 is one embodiment of a double injector layer gate stack for the NV floating plate device of Figure 17.

Figure 28 is one embodiment of the double injector layer gate stack of Figure 27.

15       Figure 29 illustrates a memory device with an array of NV memory cells according to the present invention.

### **Detailed Description of the Invention**

20       The following detailed description of the invention refers to the accompanying drawings which show, by way of illustration, specific aspects and embodiments in which the invention may be practiced. In the drawings, like numerals describe substantially similar components throughout the several views. These embodiments are described in sufficient detail to enable those skilled in the  
25       art to practice the invention. Other embodiments may be utilized and structural, logical, and electrical changes may be made without departing from the scope of the present invention. The following detailed description is, therefore, not to be taken in a limiting sense, and the scope of the present invention is defined only by the

appended claims, along with the full scope of equivalents to which such claims are entitled.

The present subject matter provides a scalable Flash/NV structure that further extends the scalability of NV technology by providing a gate stack with a high K dielectric, a charge center or charge storing medium and a least one charge injector medium. In various embodiments, the insulator stack consists of either a 4-layer element or a 5-layer element. The 4-layer element provides a tunneling medium, a charge storing medium, a charge blocking medium and a charge injector medium. The 5-layer element stack creates yet lower field programming by adding a second injector medium. According to one embodiment, the charge store layer and the charge blocking layer are co-produced as one layer, as in the case of SRI, so as to reduce the 4-layer element to a 3-layer element and to reduce the 5-layer element to a 4-layer element. The gate stack is programmable with boot-strapped circuits without a charge pump or high voltage add-on technology. According to various embodiments, the gate stack has an equivalent oxide thickness ( $t_{ox}$ ) in the range of 5 nm - 15 nm, and a programming voltage across the stack (depending on the stack thickness) as low as 4-5 V, with a programming window of ~2 V.

In addition to the power savings attributable to the reduced programming power, which is about one fourth of that for a conventional NV device, the structure of the present invention is currently believed to have greater than four orders of magnitude better endurance than possible up until this time. Additionally, the structure of the present invention has variables that are capable of being manipulated to scale the programmable voltages and power. Furthermore, since the programming voltage is significantly less than the breakdown voltage of the dielectric, these variables are capable of being manipulated to achieve faster write-erase cycles. It is believed that the write-erase cycle speed can be enhanced by greater than three orders of magnitude.

Figure 1 is a graph showing refractive index of silicon-rich silicon nitride films versus  $\text{SiH}_2\text{Cl}_2/\text{NH}_3$  flow rate ratio (R). This graph is provided herein to illustrate the known relationship between the silicon amount and the refractive index. The graph indicates that the index of refraction increases linearly with increasing silicon content. As such, the index of refraction of the films can be used as an indication of the silicon content of the films.

Figure 2 is a graph showing current density versus applied field for silicon-rich silicon nitride films having different percentages of excess silicon. The current density (J) is represented in amperes /  $\text{cm}^2$ , and log J is plotted against the electric field E (volts/cm) for  $\text{Si}_3\text{N}_4$  layers having a  $\text{SiH}_2\text{Cl}_2/\text{NH}_3$  flow rate ratio R of 0.1, 3, 5, 10, 15 and 31. This graph is provided herein to illustrate the known relationship between the amount of silicon and the conductivity of the film. The plot shows that the  $\text{Si}_3\text{N}_4$  layers having small additions of silicon (R=3 and 5) exhibit a relatively small conductivity increase over stoichiometric  $\text{Si}_3\text{N}_4$ . The plot further shows that increasing silicon content at or above R=10 substantially increases or enhances the conductivity.

Figures 3 and 4 provide graphs that illustrate the known relationship between the flatband shift and applied fields for films having varying percentages of excess silicon as represented by the  $\text{SiH}_2\text{Cl}_2/\text{NH}_3$  flow rate ratio R. Figure 3 is a graph showing flatband shift versus time at an applied field of  $4 \times 10^6$  volts/cm for silicon-rich silicon nitride films having varying percentages of excess silicon. For R=3, the flatband shift is greater than the shifts produced by films having an R of 0.1, 10 or 15. The film having an R of 10 provides a greater flatband shift than a film having an R of 15. Figure 4 is a graph showing flatband shift versus time at an applied field of  $7 \times 10^6$  volts/cm for silicon-rich silicon nitride films having varying percentages of excess silicon. The flatband shift produced by the R=3 film is even greater than that shown in Figure 3, while the shifts produced by the R=10 and R=15 films do not change as appreciably. Figures 3 and 4 are provided to illustrate the

characteristics of a charge storing medium and a more conductive charge injector medium as further explained below.

The graphs of Figures 1-4, which were described above, indicate that at low additional silicon content, silicon-rich  $\text{Si}_3\text{N}_4$  films function as a charge storing medium as they exhibit appreciably enhanced trapping characteristics (as shown by the high flatband shifts at moderate and high applied electric fields in Figures 3 and 4, respectively) without exhibiting appreciably enhanced conductivity characteristics as shown in Figure 1.

Silicon-rich silicon nitride films deposited at an R of 3 or 5 (for a refractive index of 2.10 and 2.17, respectively) will possess a charge storing function or property normally provided by a polysilicon floating gate of a EEPROM cell. In general, silicon-rich nitride films having an R greater than 0.1 and less than 10 (or, more specifically, having an index of refraction between approximately 2.10 and 2.30) will provide appreciably enhanced charge trapping or charge storing properties without providing appreciably enhanced charge conduction. This charge trapping is characteristic of a charge storing medium that can be used as a floating plate within a gate stack of a NV device.

Silicon-rich nitride films having an R greater than 10 (or, more specifically, having an index of refraction greater than 2.3) are referred to as an injector medium. Silicon nitride injectors are preferred over silicon oxide injectors because the two-phase nature of the interface is believed to provide a localized electric field distortion and an associated enhanced charge transport (high conduction). Silicon readily diffuses within silicon oxide at elevated processing temperatures, which disrupts the injection threshold by reducing the localized field distortions. However, even at higher processing temperature, silicon does not readily diffuse with  $\text{Si}_3\text{N}_4$ . A silicon-rich  $\text{Si}_3\text{N}_4$  (SRN) injector provides appreciably enhanced charge conductance without providing appreciably enhanced charge trapping over stoichiometric  $\text{Si}_3\text{N}_4$ . This is illustrated in Figures 3 and 4, which shows

progressively reduced flatband shifts for R=10 and R=15 with progressively increased conduction.

Figure 5 is a graph showing apparent dielectric constant K versus refractive index for both Silicon Rich Nitride (SRN) and Silicon Rich Oxide (SRO). The SRN and SRO plotted in this graph were provided using a Low Pressure Chemical Vapor Deposition (LPCVD) process. The SRO was fabricated at approximately 680° C, and the fabricated structure included 100 Å oxide and 150 Å SRO. The SRN was fabricated at approximately 770° C, and the fabricated structure included 45 Å oxide and 80 Å SRN. As shown in the graph, the dielectric constant of silicon is around 12. Materials with a higher K than silicon are conventionally termed a high K material, and materials with a lower K than silicon are conventionally termed a low K material. SRN that has a refractive index of 2.5 or greater and SRO that has a refractive index of 1.85 or greater have apparent dielectric constants that are greater than 12. Injector SRI includes these high K SRO and high K SRN. Charge-centered SRI includes low K SRO and low K SRN.

Figure 6 is a cross-section view of a conventional nonvolatile field effect transistor (NV FET) device such as a Flash device. The illustrated device is fabricated on a silicon substrate 612 such as a p silicon substrate or p-well in which case it is referred to as a source electrode (SE), and is separated from other devices by the isolation trenches 614. The device 610 further includes diffused regions that function as a drain region 616 and a source region 618, such as the illustrated n+ diffused regions in the p substrate. A field effect transistor (FET) channel 620 is formed in the substrate between the drain and source regions. A source contact 622 is formed to electrically couple with the source region 618, and a bit contact 624 is formed to electrically couple with the drain region 616. A floating polysilicon gate 626 is formed over the FET channel 620, and is separated from the FET channel 620 by tunnel oxide 628. A control gate 630, referred to as a program electrode (PE) for the illustrated embodiment, is formed over the floating polysilicon gate 626. An

oxide/nitride/oxide (ONO) interpoly dielectric 632 is provided around and between the PE 630 and the floating gate 626. A bit line is connected to the bit contact, and a word line is connected the PE. An oxide 633 is formed around the NV FET device.

Common dimensions for a typical NV FET device in the 0.13 to 0.15  $\mu\text{m}$  technology generations are provided below. The cell size for a NAND gate is approximately  $0.15 \mu\text{m}^2$ . The FET channel is approximately 150 nm wide. Both the floating gate and the PE are approximately 150 nm wide and about 250 nm thick. The tunnel oxide separating the floating gate from the FET channel is approximately 8 nm thick. The ONO interpoly dielectric separating the PE and the floating gate is approximately 15 nm thick. The programming voltage applied to the PE is about 16 volts, and the pulse width of a programming pulse is approximately 1 ms. The field generated across the tunnel oxide is approximately  $12 \times 10^6 \text{ V/cm}$ . The minimum program window ( $V_T(\text{"1"}) - V_T(\text{"0"})$ ) is approximately 2 V. The minimum program window is defined as the difference in the threshold voltages for a device with a stored one and a device with a stored zero. The endurance for a typical NV FET device is about  $10^5$  write/erase cycles. The power supply  $V_{DD}$  is 3.3 V.

Figure 7 illustrates the capacitive coupling for a conventional Flash device. Again, the device 710 includes a control gate or PE 730, a floating gate 726, and a substrate or SE 712. A programming voltage  $VP_1$  of 16 V is applied to the control gate. The electric field across about 8 nm of tunnel oxide 728 ( $E_{TUN.OX}$ ) is approximately  $12 \times 10^6 \text{ V/cm}$ , which reflects a coupling efficiency of about 60%. The low efficiency is attributable to the geometry and capacitor divider effects of the cell.

Figure 8 illustrates the capacitive coupling for a nonvolatile floating plate device. The device 840 includes a control gate 830 separated from a substrate by a gate stack 842. The gate stack 842 includes a tunnel insulator 844, charge centers 846 that form a floating plate capable of storing charge, and a charge blocking

dielectric 848. A programming voltage  $VP_2$  of 9.2 V is applied to the control gate 830. As there is no separate floating gate, the coupling efficiency is 100%. The average electric field  $E_{AVG}$  between the charge centers 846 and the substrate 812 is between about 6 to  $7 \times 10^6$  V/cm.

5           Figure 9 illustrates the average field enhancement due to the incorporation of a top injection layer in a gate stack for a nonvolatile floating plate device. In this illustration, the gate stack 940, which is interposed between the control gate 930 and the substrate 912, includes a tunnel layer 950, a charge blocking layer 952 that includes charge centers 946 that form a floating plate or a charge storing medium,  
10           and an injector layer 954. The gate stack dielectric is  $SiO_2$ , which has a dielectric constant of about 4. The injector layer 954 enhances the electric field by a factor of about 1.5 (1.5x). A programming voltage  $VP_3$  of 5.5 to 6.5 V is applied to the control gate 930. The resulting average electric field  $E_{AVG}$  between the charge centers 946 and the substrate 912 is reduced to about  $4 \times 10^6$  V/cm.

15           Figure 10 illustrates the average field enhancement due to the incorporation of a bottom injection layer in a gate stack for a nonvolatile floating plate device. In this illustration, the gate stack 1040, which is interposed between the control gate 1030 and the substrate 1012, includes an injector layer 1056, a tunnel layer 1050, and a charge blocking layer 1052 that includes charge centers 1046 that form a  
20           floating plate or a charge storing medium. The gate stack dielectric is  $SiO_2$ . A programming voltage  $VP_3$  of 5.5 to 6.5 V is applied to the control gate 1030. The resulting average electric field  $E_{AVG}$  between the charge centers 1046 and the substrate 1012 is reduced to about  $4 \times 10^6$  V/cm. This illustrates that the same general results are achieved whether the injector layer is on top of the gate stack or  
25           on the bottom of the gate stack. That is, the injector layer enhances the electric field by a factor of about 1.5

Figure 11 illustrates the average field enhancement due to the incorporation of both a top injection layer and a bottom injection layer in a gate stack for a

nonvolatile floating plate device. In this illustration, the gate stack 1140, which is interposed between the control gate 1130 and the substrate 1112, includes a first injector layer 1156, a tunnel layer 1150, a charge blocking layer 1152 that includes charge centers 1146 that form a floating plate or a charge storing medium, and a

5 second injector layer 1154. The gate stack dielectric is  $\text{SiO}_2$ . The use of an injector layer on the top and on the bottom of the gate stack enhances the electric field by a factor of about 1.7 (1.7x). A programming voltage  $\text{VP}_3$  of 5.5 to 6.5 V is applied to the control gate 1130. The resulting average electric field  $E_{\text{AVG}}$  between the charge centers 1146 and the substrate 1112 is reduced to about  $3.5 \times 10^6$  V/cm.

10 Figures 12, 13 and 14 correspond with the illustrations of Figures 9, 10 and 11, except that a high K dielectric such as  $\text{Al}_2\text{O}_3$  ( $K=9-10$ ) is used in the gate stack rather than  $\text{SiO}_2$  ( $K=4$ ). For example, the illustrations show a high K dielectric is used as the base dielectric for both the charge blocking layer and the tunnel layer. The high K dielectric allows the device to be scaled to smaller sizes and allows

15 smaller electric fields and programming voltages to be applied.

Figure 12 illustrates the average field enhancement due to the incorporation of a high K dielectric and a top injection layer in a gate stack for a nonvolatile floating plate device. The average field enhancement due to the  $\text{Al}_2\text{O}_3$  is approximately 1.6 greater than that for the illustration of Figure 9 that used  $\text{SiO}_2$  as

20 the gate dielectric; *i.e.* the average field enhancement is approximately 1.6 of 1.5x. A programming voltage  $\text{VP}_4$  of 3.5 to 4 V is applied to the control gate 1230. The resulting average electric field  $E_{\text{AVG}}$  between the charge centers 1246 and the substrate 1212 is, therefore, reduced to about  $2.5 \times 10^6$  V/cm.

Figure 13 illustrates the average field enhancement due to the incorporation

25 of a high K dielectric and a bottom injection layer in a gate stack for a nonvolatile floating plate device. The average field enhancement due to the  $\text{Al}_2\text{O}_3$  is approximately 1.6 times greater than that for the illustration of Figure 10 that used  $\text{SiO}_2$  as the gate dielectric; *i.e.* the average field enhancement is approximately 1.6



of 1.5x. A programming voltage  $VP_4$  of 3.5 to 4 V is applied to the control gate 1330. The resulting average electric field  $E_{AVG}$  between the charge centers 1346 and the substrate 1312 is reduced to about  $2.5 \times 10^6$  V/cm.

Figure 14 illustrates the average field enhancement due to the incorporation of a high K dielectric and both a top injection layer and a bottom injection layer in a gate stack for a nonvolatile floating plate device. The average field enhancement due to the  $Al_2O_3$  is approximately 1.6 times greater than that for the illustration of Figure 11 that used  $SiO_2$  as the gate dielectric; *i.e.* the average field enhancement is approximately 1.6 of 1.7x. A programming voltage  $VP_4$  of 3.5 to 4 V is applied to the control gate 1430. The resulting average electric field  $E_{AVG}$  between the charge centers 1446 and the substrate 1412 is reduced to about  $2.2 \times 10^6$  V/cm.

Figure 15 is a graph showing floating charge versus average programming field for programming voltages applied to a floating gate of a conventional flash device (VP1), a NV floating plate device (VP2), a NV floating device having a gate stack formed with a single injection layer (VP3), and a NV floating device having a gate stack formed with a high K dielectric and a single injection layer (VP4). The graph illustrates that a larger electric field is required to store a desired charge on a floating gate rather than on one of the floating plates. Reducing the required average programming field reduces the required programming voltage, which allows the NV device to be scaled with the other devices. Additionally, reducing the required programming voltage provides a margin with respect to the breakdown of the dielectric. A higher than required programming voltage can be applied to provide a quicker write or erase. Furthermore, for a generic write or erase speed, such reduction in programming voltage enhances endurance; *i.e.* the number of write/erase cycles.

Figure 16 is a graph showing the relationship between the log of the write/erase cycle, or endurance, and the average programming field. The graph illustrates a linear relationship between the average programming field and

endurance such that reducing the average programming field in half increases the endurance by almost four orders of magnitude, and reducing the average programming field to one fourth of an original field increases the endurance by almost eight orders of magnitude, *i.e.* by a factor of 100,000,000.

- 5           Figure 17 is one embodiment of a NV floating plate device. The device 1740 includes a gate stack 1742 that is formed on a p-type silicon substrate 1712 between two n-type diffusion regions 1716 and 1718 in the substrate. A gate 1730 is formed on the gate stack 1742. The diffusion regions in the illustrated embodiment are n+ diffusion regions, and the substrate is a p substrate. In this
- 10           embodiment, the substrate functions as a source electrode (S.E.) and the gate functions as a programming electrode (P.E.). One of ordinary skill in the art will understand, upon reading and understanding this disclosure, how to reverse the roles of the electrodes to provide the desired device operation. For example, the role of the electrodes could be reversed by placing the tunnel media either adjacent to the
- 15           silicon substrate or to the gate. The gate may be either a doped polysilicon gate such as a n+ or p+ diffused silicon, or may be a metal gate. Examples of a metal gate include TiN and WSi. One of ordinary skill in the art will understand, upon reading and understanding this disclosure, how to incorporate the gate stack into both bulk and SOI silicon based microelectronics technology, including NMOS,
- 20           PMOS and CMOS technology. The CMOS technology could either be bulk CMOS or SOI CMOS technology.

- Figure 18 is one embodiment of a single injector layer gate stack or single electron injector structure (SEIS) for the NV floating plate device of Figure 17. From the source electrode to the program electrode, the gate stack 1842 includes a
- 25           tunnel medium 1850, a combination high K charge blocking and charge storing medium 1860, and an injector medium 1854. The injector medium 1854 provides charge transfer through enhanced tunneling. The term charge storing medium connotes a medium that has charge centers that provide a charge trapping property.

Figure 19 is one embodiment of a single injector layer gate stack or single electron injector structure (SEIS) for the NV floating plate device of Figure 17. From the source electrode to the program electrode, the gate stack 1942 includes an injector medium 1956, a tunnel medium 1950, and a combination high K charge blocking and charge storing medium 1960. This illustrates that the injector medium may be used either as the top layer (near the program electrode) or as the bottom layer (near the source electrode) of the gate structure. According to one embodiment, the combination high K charge blocking and charge storing medium 1960 includes nano crystals dispersed into a high K dielectric either through doping or implantation.

In this embodiment, as is represented by the chart adjacent to the gate stack, the thickness of the tunnel  $\text{Al}_2\text{O}_3$  is 6 nm which corresponds to a  $t_{\text{OX}}$  equivalent of about 2.5 nm, the thickness of the silicon-rich  $\text{Al}_2\text{O}_3$  is 10 to 12 nm which corresponds to a  $t_{\text{OX}}$  equivalent of about 5 nm, and the thickness of the injector SRN is 5 nm which corresponds to a  $t_{\text{OX}}$  equivalent of about 1.5 nm. The total equivalent oxide thickness ( $t_{\text{OX.EQ.TOTAL}}$ ) is approximately 9 nm. A programming voltage ( $V_p$ ) of 3.6 V provides an effective programming field ( $E_p$ ) of about  $4 \times 10^6$  V/cm. A number of variables may be manipulated. An effective range for the  $\text{Al}_2\text{O}_3$  tunnel

medium is 3 to 10 nm, for the silicon-rich  $\text{Al}_2\text{O}_3$  is 6 to 30 nm, and for the SRN injector medium is 3 to 10 nm. One of ordinary skill in the art will understand, upon reading and understanding this disclosure, how to manipulate these variables to achieve the desired characteristics for the gate stack.

5           Figure 21 is one embodiment of the single injector layer gate stack of Figure 18. From the source electrode to the program electrode, the gate stack 2142 includes a  $\text{SiO}_2$  tunnel medium 2150, silicon rich  $\text{Al}_2\text{O}_3$  2160 functioning as the combination high K charge blocking and charge storing medium, and an SRN injector medium 2154.

10           In this embodiment, as is represented by the chart adjacent to the gate stack, the thickness of the tunnel  $\text{SiO}_2$  is 5 nm which corresponds to a  $t_{\text{OX}}$  of about 5 nm, the thickness of the silicon-rich  $\text{Al}_2\text{O}_3$  is 10 to 12 nm which corresponds to a  $t_{\text{OX}}$  equivalent of about 5 nm, and the thickness of the injector SRN is 5 nm which corresponds to a  $t_{\text{OX}}$  equivalent of about 1.5 nm. The total equivalent oxide  
15 thickness ( $t_{\text{OX.EQ.TOTAL}}$ ) is approximately 11.5 nm. A programming voltage ( $V_p$ ) of 4.6 V provides an effective programming field ( $E_p$ ) of about  $4 \times 10^6$  V/cm. A number of variables may be manipulated. An effective range for the  $\text{SiO}_2$  tunnel medium is 4 to 8 nm, for the silicon-rich  $\text{Al}_2\text{O}_3$  is 6 to 30 nm, and for the SRN injector medium is 3 to 10 nm. One of ordinary skill in the art will understand, upon  
20 reading and understanding this disclosure, how to manipulate these variables to achieve the desired characteristics for the gate stack.

          Figure 22 is one embodiment of a single injector layer gate stack or single electron injector structure (SEIS) for the NV floating plate device of Figure 17. From the source electrode to the program electrode, the gate stack 2242 includes a  
25 tunnel medium 2250, a charge storing medium 2264 with nano crystals, a high K charge blocking medium 2262, and an injector medium 2254. According to one embodiment, the nano crystals in the charge storing medium are disposed in a high K dielectric either through doping or implantation.

Figure 23 is one embodiment of a single injector layer gate stack or single electron injector structure (SEIS) for the NV floating plate device of Figure 17. From the source electrode to the program electrode, the gate stack 2342 includes an injector medium 2356, a tunnel medium 2350, a charge storing medium 2364 with nano crystals, and a high K charge blocking medium 2362. According to one embodiment, the nano crystals in the charge storing medium are disposed in a high K dielectric either through doping or implantation.

Figure 24 is one embodiment of the single injector layer gate stack of Figure 22. From the source electrode to the program electrode, the gate stack 2442 includes an  $\text{Al}_2\text{O}_3$  tunnel medium 2450,  $\text{Al}_2\text{O}_3$  with silicon nano crystals 2464 functioning as a charge storing medium,  $\text{Al}_2\text{O}_3$  2462 functioning as a high K charge blocking medium, and an SRN injector medium 2454.

In this embodiment, as is represented by the chart adjacent to the gate stack, the thickness of the tunnel  $\text{Al}_2\text{O}_3$  is 5 nm which corresponds to a  $t_{\text{OX}}$  equivalent of about 2.5 nm, the thickness of the  $\text{Al}_2\text{O}_3$  with silicon nano crystals is 4 nm which corresponds to a  $t_{\text{OX}}$  equivalent of about 1.6 nm, the thickness of the blocking  $\text{Al}_2\text{O}_3$  is 10 nm which corresponds to a  $t_{\text{OX}}$  equivalent of about 4.5 nm, and the thickness of the injector SRN is 5 nm which corresponds to a  $t_{\text{OX}}$  equivalent of about 1.5 nm. The total equivalent oxide thickness ( $t_{\text{OX.EQ.TOTAL}}$ ) is approximately 10 nm. A programming voltage ( $V_p$ ) of 4 V provides an effective programming field ( $E_p$ ) of about  $4 \times 10^6$  V/cm. A number of variables may be manipulated. An effective range for the  $\text{Al}_2\text{O}_3$  tunnel medium is 5 to 8 nm, for the  $\text{Al}_2\text{O}_3$  with silicon nano crystals is 3 to 5 nm, for the blocking  $\text{Al}_2\text{O}_3$  is 6 to 30 nm, and for the SRN injector medium is 3 to 10 nm. One of ordinary skill in the art will understand, upon reading and understanding this disclosure, how to manipulate these variables to achieve the desired characteristics for the gate stack.

Figure 25 is one embodiment of a double injector layer gate stack or double electron injector structure (DEIS) for the NV floating plate device of Figure 17.

From the source electrode to the program electrode, the gate stack 2542 includes an injector medium 2556, a tunnel medium 2550, a combination charge blocking and storing medium 2560, and an injector medium 2554. The use of two injector layers provides further field enhancements as provided above.

5           Figure 26 is one embodiment of the double injector layer gate stack of Figure 25. From the source electrode to the program electrode, the gate stack 2642 includes an SRN injector medium 2656, an  $\text{Al}_2\text{O}_3$  tunnel medium 2650, silicon rich  $\text{Al}_2\text{O}_3$  2660 functioning as the combination high K charge blocking and charge storing medium, and an SRN injector medium 2654.

10           In this embodiment, as is represented by the chart adjacent to the gate stack, the thickness of the first injector SRN that is deposited "NH<sub>3</sub>" or "NO" surface treated substrate is 5 nm which corresponds to a  $t_{\text{OX}}$  equivalent of about 2.5 nm, the thickness of the tunnel  $\text{Al}_2\text{O}_3$  is 5 to 6 nm which corresponds to a  $t_{\text{OX}}$  equivalent of about 2.5 nm, the thickness of the silicon-rich  $\text{Al}_2\text{O}_3$  is 10 to 12 nm which  
15           corresponds to a  $t_{\text{OX}}$  equivalent of about 5 nm, and the thickness of the second injector SRN is 5 nm which corresponds to a  $t_{\text{OX}}$  equivalent of about 1.5 nm. The total equivalent oxide thickness ( $t_{\text{OX.EQ.TOTAL}}$ ) is approximately 11.5 nm. A programming voltage ( $V_p$ ) of 3 to 3.3 V provides an effective programming field ( $E_p$ ) of about  $2.6$  to  $3.0 \times 10^6$  V/cm. A number of variables may be manipulated.  
20           An effective range for the first SRN injector media is 3 to 10 nm, for the  $\text{Al}_2\text{O}_3$  tunnel medium is 5 to 8 nm, for the silicon-rich  $\text{Al}_2\text{O}_3$  is 6 to 30 nm, and for the second SRN injector medium is 3 to 10 nm. One of ordinary skill in the art will understand, upon reading and understanding this disclosure, how to manipulate these variables to achieve the desired characteristics for the gate stack.

25           Figure 27 is one embodiment of a double injector layer gate stack or double electron injector structure (DEIS) for the NV floating plate device of Figure 17. From the source electrode to the program electrode, the gate stack 2742 includes an

injector medium 2756, a tunnel medium 2750, a charge storing medium 2764, a charge blocking medium 2762, and an injector medium 2754.

Figure 28 is one embodiment of the double injector layer gate stack of Figure 27. From the source electrode to the program electrode, the gate stack 2842 includes an SRN injector medium 2856, an  $\text{Al}_2\text{O}_3$  tunnel medium 2850,  $\text{Al}_2\text{O}_3$  with silicon nano crystals 2864 functioning as a charge storing medium,  $\text{Al}_2\text{O}_3$  functioning as a charge blocking medium 2862, and an SRN injector medium 2854.

In this embodiment, as is represented by the chart adjacent to the gate stack, the thickness of the first injector SRN that is deposited "NH<sub>3</sub>" or "NO" surface treated substrate is 5 nm which corresponds to a  $t_{\text{OX}}$  equivalent of about 2.5 nm, the thickness of the tunnel  $\text{Al}_2\text{O}_3$  is 5 to 6 nm which corresponds to a  $t_{\text{OX}}$  equivalent of about 2.5 nm, the thickness of the  $\text{Al}_2\text{O}_3$  with silicon nano crystals is 4 nm which corresponds to a  $t_{\text{OX}}$  equivalent of about 1.6 nm, the thickness of the blocking  $\text{Al}_2\text{O}_3$  is 10 nm which corresponds to a  $t_{\text{OX}}$  equivalent of about 4.5 nm, and the thickness of the injector SRN is 5 nm which corresponds to a  $t_{\text{OX}}$  equivalent of about 1.5 nm. The total equivalent oxide thickness ( $t_{\text{OX.EQ.TOTAL}}$ ) is approximately 12.6 nm. A programming voltage ( $V_p$ ) of 3.25 V provides an effective programming field ( $E_p$ ) of about  $2.6 \times 10^6$  V/cm. A number of variables may be manipulated. An effective range for the first SRN injector medium is 3 to 10 nm, for the  $\text{Al}_2\text{O}_3$  tunnel medium is 5 to 8 nm, for the  $\text{Al}_2\text{O}_3$  with silicon nano crystals is 3 to 5 nm, for the blocking  $\text{Al}_2\text{O}_3$  is 6 to 30 nm, and for the second SRN injector medium is 3 to 10 nm. One of ordinary skill in the art will understand, upon reading and understanding this disclosure, how to manipulate these variables to achieve the desired characteristics for the gate stack.

Injector SRN was provided as an example of an injector medium in the above examples. One of ordinary skill in the art will understand, upon reading and understanding this disclosure, that other materials may be used as an injector medium. These materials include silicon-rich aluminum nitride and SRO.  $\text{Al}_2\text{O}_3$  is

not an effective diffusion barrier for certain dopants like phosphorus for an n+ gate. The injector media SRN and silicon-rich aluminum nitride function as a diffusion barrier for doped polysilicon gates to prevent phosphorous, for example, from diffusing into  $\text{Al}_2\text{O}_3$ . A diffusion barrier is not needed if a metal gate is used. When  
5 injector SRN is deposited over a silicon substrate, the substrate is "NO" or " $\text{NH}_3$ " surface treated to reduce the interface states density and leakage at the silicon substrate. This surface treatment adds 1 nm to the equivalent of additional oxide thickness ( $t_{\text{OX.EQ}}$ ) of the injector SRN. This thicker  $t_{\text{OX}}$  equivalent is illustrated in Figure 25, where the top injector SRN has a  $t_{\text{OX}}$  equivalent of 1.5 nm and the bottom  
10 injector SRN deposited over "NO" or " $\text{NH}_3$ " surface treated substrate has a  $t_{\text{OX}}$  equivalent of 2.5 nm.

$\text{Al}_2\text{O}_3$  was provided as an example of a high K charge blocking medium in the above examples. One of ordinary skill in the art will understand, upon reading and understanding this disclosure, that other materials may be used as a high K  
15 charge medium. A high K charge medium is a material that has a K greater than the K of silicon. These materials include oxides, nitrides and silicates of Tantalum, Titanium, Zirconium, Hafnium and Praseodymium. Additionally, these materials may further be doped with complex high K dielectrics such as barium strontium titanate (BST), transition metal, and metal oxides such as tantalum pentoxide  
20 ( $\text{Ta}_2\text{O}_5$ ), titanium dioxide ( $\text{TiO}_2$ ), tantalum nitride (TaN), zirconium oxide ( $\text{ZrO}_2$ ), and praseodymium oxide ( $\text{Pr}_2\text{O}_3$ ).

Silicon rich  $\text{Al}_2\text{O}_3$  was provided as an example of a combination high K charge blocking and charge storing medium in the above examples. One of ordinary skill in the art will understand, upon reading and understanding this disclosure, that  
25 other materials may be used as a combination high K charge blocking and charge storing medium. These materials include any of the high K charge blocking media provided above with nano crystals dispersed through the media. Various



embodiments include a high K charge blocking media with silicon nano crystals, gold nano crystals, tungsten nano crystals, and/or silicided tungsten nano crystals.

In the examples provided above it was indicated that the nano crystals were dispersed into a high K dielectric either through doping or implantation. One of  
5 ordinary skill in the art will understand, upon reading and understanding this disclosure, that the nano crystals may be distributed using a number of techniques, including simultaneous sputtering, implantation, chemical vapor deposition, atomic layer deposition (ALD) and molecular beam epitaxy (MBE).

$\text{Al}_2\text{O}_3$  and  $\text{SiO}_2$  were provided as examples of a tunnel medium in the above  
10 examples. These materials may be interchanged with each other.  $\text{Al}_2\text{O}_3$  has a higher dielectric constant, whereas  $\text{SiO}_2$  is easier to fabricate.

Figure 29 illustrates a memory device with an array of NV memory cells according to the present invention. The memory device 2970 includes an array 2972  
15 of NV memory cells as described above. A grid of row conductors 2974 and column conductors 2976 are used to selectively write/erase a memory cell 2978. Additionally, the memory device 2970 includes power circuitry 2980, row select circuitry 2982 and column select circuitry 2984. The row select circuitry and column select circuitry cooperate with each other to select a memory cell to be  
20 written or erased using power provided by the power circuitry. Input/output circuitry and pads, not shown, defines the inputs and outputs of such device. According to various embodiments, the memory device is used in a number of nonvolatile multi-threshold FET devices such as PROM, FLASH, EPROM, EEPROM, and antifuse devices.

Furthermore, one of ordinary skill in the art will understand, upon reading  
25 and comprehending this disclosure, how to incorporate NV memory cells according to the present invention into a larger electronic system. Such an electronic system includes a processor or arithmetic/logic unit (ALU), a control unit, a memory device unit and an input/output (I/O) device. Generally, such an electronic system will

5  
10

15

### CONCLUSION

The present subject matter provides a gate stack for a nonvolatile multi-threshold FET device that promotes low power, low programming voltage for write and erase cycles, and improved endurance. A higher capacitive coupling efficiency is achieved by replacing floating gate technology with floating plate (charge center) technology, and using a high K dielectric. Furthermore, an injector medium enhances the tunneling effect. The gate stack is capable of being used in applications which are compatible and scalable with power supply and lithography scaling. Additionally, the gate stack is capable of being used in devices and circuits that readily integrate with general fixed threshold memory and logic devices, circuits and functions. Due the lower programming voltage and electric field, which is considerably less than the breakdown of the dielectric, the gate stack promotes faster writing and erasing capabilities.

Although specific embodiments have been illustrated and described herein, it will be appreciated by those of ordinary skill in the art that any arrangement which is calculated to achieve the same purpose may be substituted for the specific embodiment shown. This application is intended to cover any adaptations or variations of the present invention. It is to be understood that the above description is intended to be illustrative, and not restrictive. Combinations of the above embodiments, and other embodiments will be apparent to those of skill in the art upon reviewing the above description. The scope of the invention includes any other applications in which the above structures and fabrication methods are used. The scope of the invention should be determined with reference to the appended claims, along with the full scope of equivalents to which such claims are entitled.

25